图书情报工作
第 63 卷 第 11 期 2019 年 6 月

社会化问答社区用户生成答案质量自动化评价研究*
——以“知乎”为例

郭顺利[1] 张向先[2] 陶兴[2] 张莉曼[2]
[1] 曲阜师范大学传媒学院 日照 276826 [2] 吉林大学管理学院 长春 130022

摘要：[目的/意义]旨在构建社会化问答社区用户生成答案质量评价指标体系，实现面向用户需求的答案质量自动化评价和筛选，提高社会化问答社区知识服务质量。[方法/过程]引入社会情感特征和用户特征，运用因子分析和结构方程实证构建用户生成答案质量评价指标体系。基于 GA-BP 神经网络模型设计答案质量自动化评价方法。最后，选取知乎网站数据对用户生成答案质量评价指标体系和自动化评价方法进行应用研究。[结果/结论]构建包含答案文本特征、回答者特征、时效特征、用户特征、社会情感特征 5 个维度的评价指标体系。实验分析发现基于 GA-BP 神经网络的答案质量自动化评价方法相比于其他方法准确率较高、平均误差低，具有可行性和有效性，能够进一步应用和推广实践。

关键词：社会化问答社区 用户生成答案 质量评价 用户需求
分类号：G203
DOI：10.13266/j.issn.0252-3116.2019.11.013

近年来社会化问答社区迅速发展，注册用户数量均呈现指数式增长。以知乎为例，自 2013 年向公众开放注册，截至 2018 年 11 月底，知乎官方宣布注册用户已经超过 2.2 亿，其问题数超过 3 000 万，回答数超过 1.3 亿。社会化问答社区已发展成为多元化、机制完善的大型知识分享平台，成为人们日常获取信息与知识的重要途径。然而，社会化问答社区具有社会化和开放性的特点，平台上的问题和答案以用户参与生成为主。任何用户都可以随意地提问和回答问题，这使得用户生成答案的质量良莠不齐。而且提问用户受到自身经验和认知局限，其所采纳的答案未必是最佳答案，有的甚至是恶意的广告或虚假信息，从而使得用户在社会化问答社区搜寻、鉴别和获取知识等方面付出了大量的时间和精力，出现“知识过载和迷航”现象，降低了用户知识搜寻和获取的效率，难以获得良好的用户体验。另外，随着社会化问答社区用户规模扩大，用户生成问题和答案数量也越来越多，通过人工方式进行答案质量评价变得困难而且效率低下，仅仅依靠人工审核或标注难以解决当前社会化问答社区面临的问答质量问题。因此，用户生成答案质量自动化评价成为社会化问答社区运营亟需解决的问题。

1 国内外相关研究现状

1.1 问答社区答案质量评价特征选取

国内外学者尝试从数据质量框架、信息质量评价标准、外部线索等不同角度选取用户生成答案质量评价特征，验证不同特征对答案质量的影响，加入诸如情感、时效等特征维度，并针对不同问答平台进行应用研究，但是仍然没有形成统一的评价标准。

国外方面，S. Kim 等[1]研究发现 Yahoo! Answers 用户选取和采纳最佳答案时会考虑社会性情感、内容及效用相关的评价标准，并且不同话题的评价标准也存在差异；D. Ishikawa 等[2]构建了包括回答者经验、证据来源、礼貌程度、详细程度、意见、相关性、具体化程度、全面性等 12 个维度的问答社区答案质量评价指标体系。S. Oh 等[3]选取信息准确性、完整性、相关性、来源可靠性、回答者同情心、客观性、可读性、礼貌、自信、回答者的努力 10 个指标作为衡量答案质量的评价标准，对比分析不同职业人员对问答社区答案质量的评估差异。P. Fichman[4]从准确性、完整性、可证实性

* 本文系吉林大学研究生创新基金资助项目“移动商务用户在线评论的隐式意见挖掘及可视化研究”(项目编号:2017082)研究成果之一。
作者简介：郭顺利(ORCID:0000-0002-3155-9937)，讲师，博士，E-mail:guosl777@163.com；张向先(ORCID:0000-0003-3186-2677)，教授，博士，博士生导师；陶兴(ORCID:0000-0003-0480-4201)，博士研究生；张莉曼(ORCID:0000-0002-0770-3708)，博士研究生。
收稿日期:2018-08-05 修回日期:2018-11-18 本文起止页码:118-130 本文责任编辑:刘远颖

3 个方面对问答社区答案质量评价，发现部分非主流问答网站的答案质量也很高，问题回答质量与问答社区平台自身关系较小。A. Y. K. Chua 等[5]研究了回答速度与答案质量之间的关系，发现不同类型问题的回答质量和回答速度之间存在显著差异，最优质的答案比最快的答案有更好的整体回答质量。

国内方面，学者们主要从答案的文本、非文本等不同角度选取特征指标构建答案质量评价指标体系。孙晓宁等[6]从内容质量、情境质量、来源质量和情感质量 4 个维度，实证构建了社会化搜索答案质量评价模型。李翔宇等[7]结合专家评分法及三角模糊加权平均 G1 法，构建了 SQA 平台答案质量评测指标体系，并证实了答案质量评测指标体系的科学性。张煜轩[8]结合线索理论基于用户视角，发现信息利用线索、信息认同线索、信息举报线索、信息否定线索、信息能力线索、信息表象线索、系统推荐线索 7 类外部线索对用户感知判断社会化问答社区信息质量产生影响，提出了基于外部线索的社会化问答平台的信息质量感知模型。姜雯等[9]将情感特征引入在线问答社区信息质量评价，从文本特征、用户特征、时效特征、情感特征 4 个维度评价在线问答社区信息质量；袁红等[10]从信息质量定义出发构建了回答形式、回答内容和回答效用 3 个维度的问答社区答案质量评价指标体系。孔维泽等[11]从基于文本特征、时序特征、链接特征、问题粒度特征和百度知道社区用户特征角度对问答社区答案质量进行评价。罗毅等[12]引入新的 RIPA 理论，认为用户生成内容的完整性、专业性和权威性 3 个指标是影响社会问答平台答案质量的关键因素。

1.2 问答社区答案质量评价方法研究

目前国内外学者一般将答案质量评价视为基于机器学习的分类问题[13]，选取机器学习方法应用于问答社区答案质量评价，例如最大熵、支持向量机、决策树、随机森林、逻辑回归、神经网络等。部分学者采用层次分析、模糊综合评价等传统评价方法，也有部分学者基于构建的评价指标体系进行人工性标注评价，采用人工评价和自动化评价相结合的方法。国外研究方面，部分学者为提高最佳答案的发现和预测精准性，将答案质量评价视为分类问题，通过改进分类算法提高最佳答案发现和预测的精准性和召回率。例如：J. Jeon 等[14]提出基于非文本特征的问答社区答案质量预测方法，实证研究发现比基于基础特征的问答社区答案质量预测具有显著的改进。C. Shah 等[15]以 Yahoo! Answers 为例，首先采用人工标注评价给定问题的答案质量，通过提取问题、答案和用户的各种特征训练分类器进行最佳答案选取研究。

国内方面，李晨等[16]基于给定的问答质量判定标准，通过提取文本和非文本两类特征集，利用机器学习算法设计和实现了基于特征集的问答质量分类器。王伟等[17]将结构化特征、文本特征、用户社交属性引入中文问答社区答案质量评价特征体系，然后选取逻辑回归、支持向量机和随机森林 3 种评价方法，结合新设计的 3 个方面特征和经典的文本特征、链接特征，对高质量和非高质量的回答进行分类研究。崔敏君等[18]基于问题类型提取文本、非文本、语言翻译性、答案中的链接数 4 类特征，采用逻辑回归算法对各类型问题的答案质量进行评价。胡海峰等[19]从答案的文本信息和非文本信息的特征表示与融合两方面入手，针对社区问答系统用户生成答案质量评价方法开展研究。

1.3 研究述评

通过梳理已有的研究成果，不难发现当前国内外主要是采用单一特征或多个特征指标组合的方式构建用户生成答案质量评价指标体系，但是构建的评价指标体系存在不够全面、没有统一标准、部分指标具有主观性和模糊性、难以进行量化和判断等问题。很少有研究考虑用户社会情感特征对答案质量评价的影响，也没有考虑用户需求、兴趣爱好、认知水平等个体差异性特征，缺乏形成面向用户需求的个性化评价指标体系。学者们将答案质量评价看作是机器学习分类问题，运用 SVM、随机梯度增强、决策树、最大熵、逻辑回归、贝叶斯、J48 等方法，均取得了良好的实验效果。虽然目前存在大量的针对答案质量自动化评价研究，但是很少有学者采用神经网络方法进行评价，没有对比其与其他方法有效性和准确性上的差异。

鉴于此，本文拟结合前人的研究成果，从用户需求角度构建用户生成答案质量自动化评价指标体系，试图解决评价指标模糊化、不够全面、缺乏个性化等问题，并将答案质量自动化评价看作是机器学习问题，选取了机器学习中典型方法遗传算法优化 BP 神经网络模型，基于本文构建的用户生成答案质量自动化评价指标体系开展实证应用研究，提出一种社会化问答社区用户生成答案自动化评价方法。

2 用户生成答案质量评价指标体系构建

2.1 评价指标的初步选取

本研究参照文献[13]在分析答案质量评价指标的基础上，认为用户在评价答案质量过程中受到多方

面因素的影响，一般情况下需要考虑答案文本内容质量、回答者质量、时效性，大部分研究也证实了这 3 类特征对答案质量的影响。然而，社会化问答社区作为开放的社交类网站，用户在筛选和评价答案过程中，也会考虑其他用户对于答案质量的评价情况（诸如：点赞、转发、评论等），容易受到周围的人际关系、社区意见领袖、交流互动等因素影响，问答社区中意见领袖能够影响其他用户的认知，他们的答案能够获得较多粉丝的支持和赞同[17]。而且，回答者的回答情感态度和积极程度也会影响用户采纳答案。所以，本文将用户对于答案的社会情感态度特征引入答案质量评价。另外，社会化问答社区不同的用户受到认知、需求和兴趣爱好等自身特征影响，对于答案质量评价拥有不同的标准和要求。因此，答案质量评价过程中还需要考虑用户自身的特征，使得筛选的答案更满足用户个性化需求。

所以，本文将用户社会情感和用户自身特征引入答案质量评价，将用户生成答案质量评价指标分为 5 个维度，分别是答案文本特征维度、回答者特征维度、时效性维度、用户特征维度、社会情感维度。然后通过阅读和综述大量的有关于信息质量评价文献，并在信息系统成功模型、使用与满足理论、数据质量框架[20]等理论研究的基础上，初步选取了 24 个评价指标，如表 1 所示：

表 1　用户生成答案质量自动化评价指标初步筛选结果

维度	指标	解释及说明	主要参考文献来源
答案文本特征	文本长度	答案文本包含的字符数。答案文本的长度越长，答案越丰富和完整	[13][20-26]
	关键词数量	答案文本中包含的关键词数量	[17][22][24]
	句子数量	答案文本中包含的句子数量	[24]
	停用词数	答案文本中包含的通用词数量，停用词数量越少，质量越高	[22]
	问题与答案耦合度	提问问题与答案之间的重叠部分，文本长度之比	[23][25-28]
	外部链接数量	答案文本中包含的超链接的数量	[29]
	段落数	答案文本的段落数	[27]
	问题答案长度比	问题长度与答案长度的比值	[23]
回答者特征	最佳答案数量	回答者的所有答案中被选为最佳答案的数量	[22-23][29-30]
	回答问题数量	回答者的所有回答的数量，表明回答者的经验和参与积极性	[15][22-23][29-30]
	用户权威性	回答者的社区等级（积分），表明专业程度和影响力	[13][29-30][31]
	提问数量	回答者提问问题的数量	[29-30]
时效性	答案的相对回答次序	答案在所有答案中的相对位置	[15][22-23][26]
	答案与问题生成间隔时间	回答时间与提问时间的间距	[13][23][26][32]
用户特征	用户学历水平	提问者的专业水平和学历程度	[15][25][29][33]
	用户提问数量	提问者以往提问问题的数量	[15][25][29][33]
	用户偏好与答案耦合度	用户的习惯，个人偏好信息需求与答案的关联性	[15][25][29][33]
	用户等级	提问者的权威性和影响力	[13][15][21][29][33]
社会情感	情感特征词数量	答案文本中包含的情感词的数量	[13]
	回答者情感态度	答案文本呈现出的回答者情感态度倾向性	[13][24]
	赞同数量	答案被赞同/支持的数量	[16][22-23]
	反对数量	答案被反对/踩的数量	[22-23]
	评论互动数量	答案被评论的数量	[15-16][29][32]
	关注关系	回答者与提问者的好友关系	自设

初步选取社会化问答社区用户生成答案质量评价指标后，笔者采用专家访谈方法修正相关表述，重点从指标的合理性、完整性两个角度听取专家的意见，探讨评价指标维度划分和选取是否合理，指标名称是否恰当，是否存在模糊性、难以测量等问题，消除指标的歧义和模糊性，初步实现指标的规范化筛选。最后依据专家的建议和反馈，形成用户生成答案质量自动化评价指标，见图 1。具体的修正如下：

（1）删除回答者特征维度的“回答者提问问题数量”，因为回答者提问问题的数量体现了回答者需求，不能体现回答者生成答案的能力和经验水平，对于回答者生成答案质量影响不够明显。删除答案文本特征中的“停用词数量”“段落数”“答案与问题的耦合度”3 个指标，因为答案文本的停用词数应该是答案文本长

度减去关键词数量，指标之间存在重复性。答案文本特征中"段落数"不能够体现答案质量，对答案质量的影响较小。另外，"答案与问题耦合度"和"用户偏好与答案耦合度"重复，用户提问问题是用户需求与偏好的体现，所以删除答案文本特征中的"停用词数量""段落数""问题与答案的耦合度"3 个指标。删除用户维度的"用户学历水平"指标，因为用户学历水平对用户判断答案文本质量影响较小。

（2）评价指标因素的补充。答案文本维度增加"图片或动画数量"指标，由于移动互联网环境下，很多用户乐于通过图片或动画来理解或掌握知识，并且图片或动画包含的信息量大，能够使用户易于理解和掌握答案内容。所以，用户生成答案中的图片或动画的数量影响答案质量。回答者特征维度补充"专业领域与问题的匹配度"指标，体现回答者的专业程度和对于问题的了解熟悉程度。

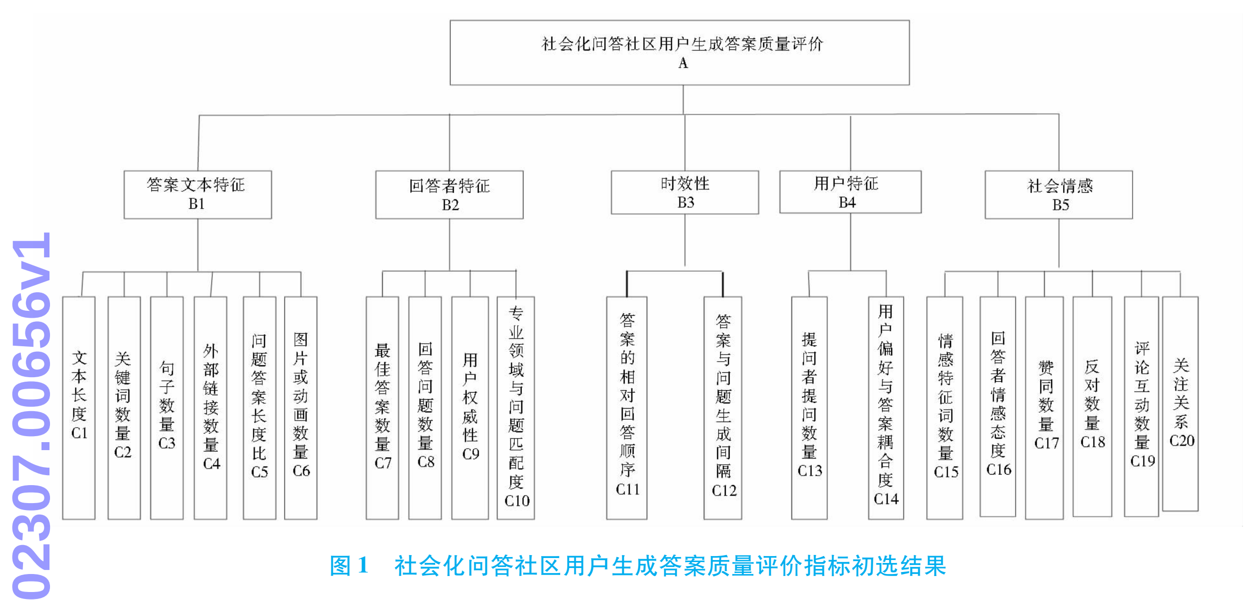


图 1　社会化问答社区用户生成答案质量评价指标初选结果

2.2　用户生成答案质量评价指标体系实证分析

2.2.1　问卷设计和发放　调查问卷主要是对社会化问答社区用户生成答案质量的测量，采用陈述句的方式表达各个评价指标对于答案质量测量的可行性和合理性。通过网络和实地发放的形式，总计回收调查问卷 610 份，其中有效问卷 580 份，将获取的调查问卷随机分为两部分（每部分样本数量 290 份），分别用于 EFA（探索性因子分析）和 CFA（验证性因子分析）。

通过对获取的调查样本数据信度检验发现 5 个维度的 Cronbach's 的值均大于 0.8，总体样本的信度为 0.846，说明调查样本获取的数据信度好，具有很好的可靠性。但是删除指标"问题与答案长度比值 C5""关注关系 C20"后，答案文本特征维度 B1 和社会情感维度 B5 的信度明显提高，问卷总体信度也会明显提高，说明指标问题与答案长度比值 C5 和"关注关系 C20"没有通过信度检验，应该予以删除。然后进行 KMO 和 Bartlett 球形检验，分析发现，Bartlett 球形检验近似卡方显著性，说明相关矩阵中存在公因子，样本整体的效度很好，适合进一步做因子分析。

2.2.2　探索性因子分析　本文采用主成分分析法进行 EFA，发现抽取到 5 个公因子时累积方差贡献率达到了 55.051%。运用最大方差进行因子旋转，迭代 10 次后收敛循环，得到旋转因子矩阵。结果如表 2 所示：

表 2　旋转因子矩阵

变量	公因子				
	1	2	3	4	5
C1	**0.617**	0.028	0.166	0.026	0.067
C2	**0.593**	0.163	0.059	0.029	0.08
C3	**0.693**	0.126	－0.121	0.056	0.179
C4	**0.581**	0.124	0.041	0.196	－0.097
C6	**0.633**	－0.057	0.2	0.035	－0.072
C7	0.044	0.112	**0.732**	0.037	0.184
C8	0.182	0.058	**0.71**	0.17	0.078
C9	0.053	0.393	**0.701**	0.089	0.007
C10	0.161	0.265	**0.584**	0.154	0.112
C11	0.063	0.274	0.19	0.168	**0.587**
C12	0.199	0.372	0.009	0.116	**0.698**
C13	0.12	**0.804**	0.271	0.112	0.054
C14	0.131	**0.763**	0.212	0.206	－0.001
C15	－0.023	0.5	0.105	**0.472**	**0.496**
C16	0.179	0.323	0.072	**0.698**	0.038
C17	－0.064	－0.043	0.389	**0.487**	0.396
C18	－0.029	－0.137	0.147	**0.653**	0.213
C19	0.247	－0.009	0.124	**0.694**	0.235

从表2中可以得出,评价指标变量"情感特征词数量C15"在公因子2、公因子4、公因子5上的载荷因子很接近,差别不是很明显,效度很差,应该予以删除。公因子1解释了C1、C2、C3、C4、C6共5项指标变量,对应了答案文本特征维度的全部指标;公因子2解释了C13、C14共2项指标变量,对应用户特征维度的全部指标;公因子3解释了C7、C8、C9、C10共4项指标变量,对应了回答者特征维度的全部指标;公因子4解释了C16、C17、C18、C19,对应社会情感维度的除C15之外的4个指标;公因子5仅包含了C11、C12这两个指标变量,对应时效性维度指标。这与我们前边初步假设的维度一致,说明本文将用户生成答案质量的评价指标分为5个维度较为合理,后续笔者将进一步结合CFA的检验结果进行修正。

2.2.3 验证性因子分析 采用结构方程模型软件AMOS17.0软件进行验证性因子分析(CFA),利用另外一部分样本数据(290份)进一步检验指标的有效性,共设置了17个观察变量、5个潜在变量、17个残差变量。采用最大似然估计方法,观测变量与其对应潜在变量之间的载荷关系系数估计,如表3所示:

表3 观测变量与其对应潜在变量之间的载荷关系系数估计

对应关系	非标准化值	标准化估计值	S. E	C. R.	P	是否支持
C1←答案文本特征	1	0.560				支持
C2←答案文本特征	0.970	0.507	0.152	6.394	***	支持
C3←答案文本特征	1.289	0.638	0.174	7.401	***	支持
C4←答案文本特征	1.425	0.757	0.180	7.896	***	支持
C6←答案文本特征	1.012	0.512	0.157	6.444	***	支持
C7←回答者特征	1	0.589				支持
C8←回答者特征	0.946	0.616	0.122	7.753	***	支持
C9←回答者特征	1.185	0.751	0.137	8.646	***	支持
C10←回答者特征	0.776	0.641	0.097	7.962	***	支持
C11←时效性	1	0.727				支持
C12←时效性	1.027	0.707	0.138	7.449	***	支持
C13←提问者特征	1	0.875			***	支持
C14←提问者特征	0.917	0.831	0.080	11.466	***	支持
C16←社会情感	1	0.607				支持
C17←社会情感	1.324	0.590	0.186	7.113	***	支持
C18←社会情感	0.854	0.383	0.162	5.259	***	不支持
C19←社会情感	1.279	0.674	0.168	7.609	***	支持

根据一般性的经验法则,如果C. R.绝对值大于2.58,表示模型的参数估计值达到了0.01显著水平,路径系数获得数据的支持;当P值小于0.001时,显示"***",表示模型达到了0.001的显著水平[34]。从表3可以得出,评价指标体系显著性检验中"反对数量C18"的标准化载荷因子估计值小于0.5,说明该指标没有通过效度检验,应该予以删除。然后,利用AMOS提供的模型拟合度评价指标来评价构建的评价指标体系的合理性,根据各指标的检验标准,发现相关指标检验结果均在可接受的范围之内,总体上构建的评价指标体系基本达到了检验的要求。当删除观测变量C18后,发现模型的绝对适配度指标χ^2值由186.125减少到150.568,*CMIN/DF*值由1.708减少到1.602,说明模型的绝对适配度性能提高,所以,更加进一步证实删除指标C18,能够提高构建的评价指标体系合理性。

2.3 评价指标的修正和确立

采用探索性因子分析和验证性因子分析等实证分析后,综合考虑EFA和CFA的检验结果,由于答案文本维度的"答案与文本长度比值C5"获取数据未通过信度检验,而且与"文本长度C1"之间存在一定的重复性,所以将其删除;"关注关系C20"也没有通过信度检验,所以也将其删除;对社会情感维度的"情感特征词数量C15"进行主成分分析时,载荷因子在多个公因子上的差别不是很明显,效度很差,而且与"回答者情感态度倾向"之间存在相关性,所以应该予以删除。另外,对指标"反对数量C18"进行载荷系数检验时,其载荷系数小于0.5不符合显著性检验标准,同时删除后模型整体的适配度和同维度指标的载荷系数得到明显提升,所以将其删除。综上所述,最终选取的社会化问答社区用户生成答案质量自动化评价指标包括5个维度、16个指标,如图2所示:

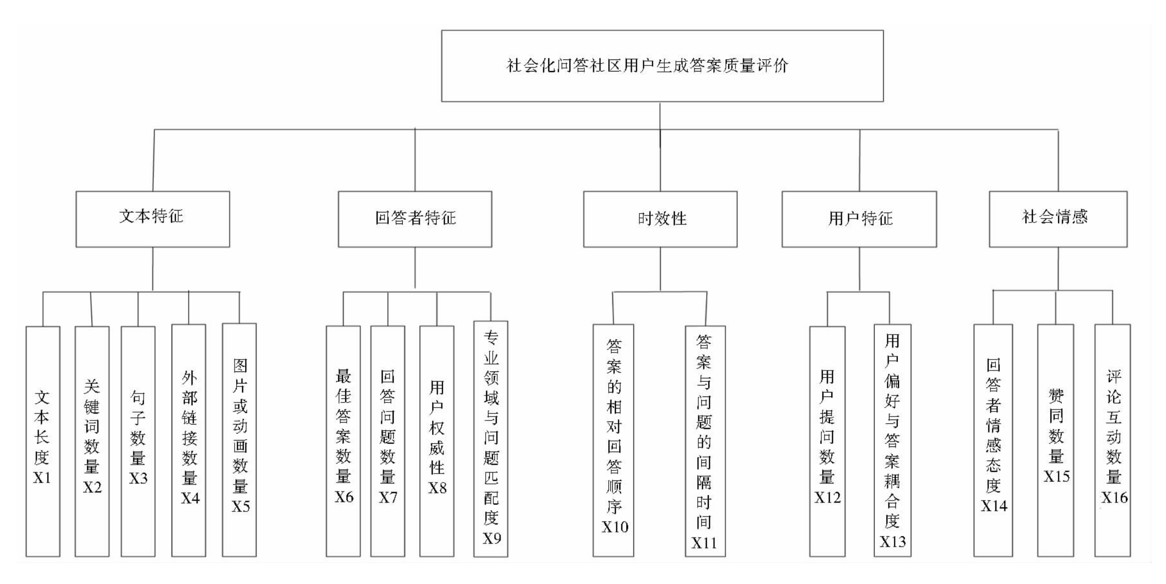


图2　社会化问答社区用户生成答案质量自动化评价指标体系

3　基于 GA-BP 神经网络的用户生成答案质量自动化评价

3.1　评价指标的获取方式和量化

社会化问答社区用户生成答案质量自动化评价是计算机运用算法程序自动化实现评价，也需要实现评价指标的自动量化。运用 GooSeeker 软件爬虫程序直接采集数据和文本信息，借助 jieba 分词工具、Hownet 情感词典、文本处理技术等工具方法，编写 python 程序和 Matlab 程序实现指标的统计和量化。

（1）文本特征维度指标获取方式和量化。包括：①文本长度。文本长度可以直接采用答案文本的字符数进行量化。一定的阈值范围内，通常认为答案文本长度越长，所包含的有用信息越多，越能够更好地满足用户知识需求，答案质量越高。②关键词数量。关键词数量可以采用答案文本中含有的除停用词外的词频统计数值进行量化。将总词频数量减去停用词词频数量得到关键词数量。③句子数量。句子数量采用文本中出现句号、问号等表示句子结束符号的次数进行量化。④外部链接数量。外部链接是指答案文本中出现的参考来源和答案扩展链接，可以采用答案文本中出现的超链接的数量直接统计量化。⑤图片或动画数量。图片或动画数量直接通过答案文本中图片或动画数量统计得到。

（2）回答者特征维度指标获取方式和量化。包括：①最佳回答数量。最佳回答数量采用问答者回答所有答案中被采纳为最佳答案的数值进行量化，也可以采用被采纳为最佳答案率进行量化。知乎问答社区回答者被知乎日报或知乎圆桌收录的回答数量，可用于量化最佳答案数量。②回答问题数量。回答问题数量采用用户回答的所有问题数量进行量化。③用户权威性。用户权威性采用回答者的用户等级或积分直接进行量化。用户等级或积分越高，表示用户获得问答社区认可度越高，影响力和权威性越大的可能性越高。④专业领域与问题匹配程度。如果专业领域与问题领域匹配，则为 1，不匹配的话为 0。

（3）时效性维度指标获取方式和量化。包括：答案的相对回答顺序是指同一问题下，答案按照回答时间进行排序，当前答案在所有答案回答时间的顺序位置。采用以下的方式量化：

$$答案相对回答顺序 = \frac{答案回答的时间顺序}{所有答案个数}$$

答案与问题的生成间隔可以采用回答日期与提问日期之间的天数差值来进行数值量化。同时为了避免数值过大造成偏差，运用分组的方法进行消除。经过问卷调查和访谈后，天数差值的取值范围和量化 10 分制数值如表 4 所示：

表 4　答案与问题的生成间隔时间数值量化对应

天数差值	量化数值	天数差值	量化数值
[0－1)	10	[14－30)	5
[1－3)	9	[30－90)	4
[3－5)	8	[90－180)	3
[5－7)	7	[180－360)	2
[7－14)	6	360 天及以上	1

（4）用户维度指标获取方式和量化。包括：①提问者提问数量。一般社会化问答社区用户基本信息中

都包括提问者的提问问题数量,直接采用爬取数值方式进行量化。例如:知乎用户基本信息中包括"提问数"这个信息,可以直接根据提问数的数值进行量化。②用户偏好与答案耦合度。运用答案文本和用户偏好两个向量之间的相似度大小来进行量化,本文认为问题是用户知识需求的最直接体现,可以采用问题与答案文本之间的相似度进行度量。

(5)社会情感维度指标获取方式和量化。包括:赞同数量、评论互动关系可以通过爬取数据进行直接量化;回答者情感态度包括正向情感、负向情感、中立 3 种极性,采用答案文本中出现的情感词数量来量化回答者情感态度。以情感基础词典为标准对答案文本进行情感特征词的数量统计,以实际统计词语数量为量化数值;采用答案下方评论数量化评论互动数量。

3.2 基于 GA-BP 神经网络的用户生成答案质量评价方法

BP 神经网络是一种利用误差反向传播训练算法的神经网络,也是应用最广泛的人工神经网络算法,包括输入层、隐层和输出层 3 层结构。标准 BP 神经网络通过有监督的学习方式进行学习和训练,采用误差函数按梯度下降的方法学习,使网络的实际输出值和期望输出值之间的均方误差最小[35]。虽然 BP 神经网络已经被广泛地应用于各个领域,但是存在易陷入局部极小值、不能保证收敛到全局最小点、收敛速度慢、训练时间过长等问题。然而,遗传算法(Genetic Algorithm, GA)用概率化的寻优方法,自动获取和指导优化搜索空间,自适应地调整搜索方向,不需要确定的规则,具有很强的全局搜索能力和全局优化性能[36]。遗传算法具有较好的全局搜索能力,容易得到全局最优解,很好地克服 BP 算法局部最优缺陷,且能够优化 BP 神经网络初始权重和阈值。因此,选用遗传算法优化 BP 神经网络(简称"GA-BP 神经网络"),能够使得 BP 神经网络的收敛速度加快,提高网络的预测精度和稳定性。

社会化问答社区用户生成答案质量受到 5 个维度 16 个特征因素的影响,其答案质量自动化评价结果很难用数学解析式来表示,属于典型的非线性问题。然而 BP 网络作为多层前馈型网络,具有强大的非线性映射能力,它能够模拟分析 5 个维度 16 个评价指标因素之间的非线性关系,可以实现非线性分类和预测,通过反复的学习训练之后可以充分地逼近任何较为复杂的非线性关系。另外,GA-BP 神经网络算法已经被广泛的应用于其它领域,并取得了丰硕的研究成果,拥有较好的理论和实践基础,能够使得社会化问答社区用户生成答案质量评价方法更具有客观性和合理性。因此,本研究采用遗传算法改进 BP 神经网络来实现社会化问答社区用户生成答案质量自动化评价。训练 BP 神经网络前先用遗传算法对 BP 神经网络的初始权值和阈值进行寻优,缩小搜索范围之后,再利用 BP 神经网络算法进行自动化评价。

基于 GA-BP 神经网络的社会化问答社区用户生成答案质量评价过程如图 3 所示:

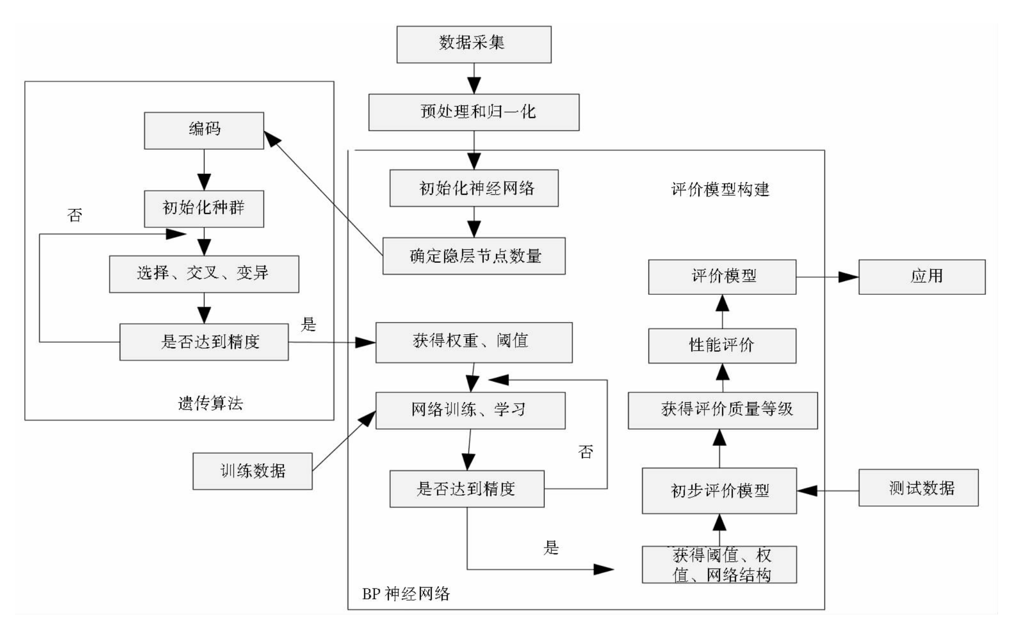


图 3 社会化问答社区用户生成答案质量自动化评价过程

3.2.1 指标特征提取量化和归一化处理 首先，运用自动爬虫采集软件GooSeeker软件自动化爬取数据，采用3.1小节指标量化方法提取各个评价指标特征。由于提取到的样本数据中各个指标量化值拥有不同的数量级，而且相互之间存在差距。在运用GA-BP神经网络计算与评价时，如果数据之间的差别过大，容易导致网络的权重也有同样数量级的差别，致使构建的网络非常“敏感”。为了确保BP神经网络的训练速度和精度，避免因为数据过大或者过小造成误差，需要将采集到的数据进行归一化处理。本文采用S型函数作为激活函数，S型激活函数的值域限制在[-1,1]，所以采集到样本数据需要归一化到区间[-1,1]，采用premnmx函数对提取到的样本数据进行归一化处理，如公式(1)、公式(2)所示：

$$PN=\frac{2(p-\min p)}{\max p-\min p}-1 \qquad 公式(1)$$

$$TN=\frac{2(t-\min t)}{\max t-\min t}-1 \qquad 公式(2)$$

公式(1)、公式(2)中p和t分别表示输入数据和输出数据值，$\min p$和$\max p$分别表示输入数据的最小值和最大值，$\min t$和$\max t$分别表示输出数据的最小值和最大值。

3.2.2 初始化BP神经网络

(1)输入层、输出层确定。Kolmogorov[37]定理证明BP神经网络中采用1层隐层网络就能够以任意精度去逼近任意映射关系。因此，为了简化模型的复杂度和提高GA-BP神经网络的学习速度与效率，本研究将社会化问答社区用户生成答案质量评价模型网络结构设置为3层，仅包括1层隐层。社会化问答社区用户生成答案质量评价的5个维度16个指标作为GA-BP神经网络的输入层，即GA-BP神经网络的输入层神经元个数为16个。输出层输出的结果反映社会化问答社区用户生成答案质量的高低，所以输出层的神经元个数为1。

部分研究中采用人工标注方式对答案质量等级进行标注并作为输出变量，将答案质量等级分为非常低、低、一般、高、非常高5个等级，但是人工标注可能与用户真正需求存在差异。为了体现出用户信息需求的差异性，本文将用户选取的最佳答案定义为最高级，然后计算其余答案文本与最佳答案之间的相似度，依据相似度将答案质量进行等级划分，见表5。如果没有最佳答案，则选择赞同/支持票数最多的答案为最佳答案。

表5 答案文本质量等级划分

答案文本之间相似度	答案质量等级	量化分数(相似度*10)
[0,0.2)	非常低	[0,2)
[0.2,0.4)	低	[2,4)
[0.4,0.6)	一般	[4,6)
[0.6,0.8)	高	[6,8)
[0.8,1.0]	非常高	[8,10)

(2)隐层节点数量确定。本文选取试凑法确定BP神经网络的隐层节点数量。在BP神经网络中其它参数值保持不变的情况下，使用同一样本集进行训练，通过调整隐层神经节点的数目重复测试，选取MSE取最小值的节点数量作为最佳隐层神经元节点的数目。采用公式(3)计算方法，得到一个粗略的估计值作为初始值，再用试凑法确定最佳隐层节点数。

$$n_l=\sqrt{n+m}+a \qquad 公式(3)$$

公式(3)中，n_l为隐层节点个数，n为输入层节点数，m为输出层节点数，a为1-10之间的常数。

(3)初始化函数设定。BP神经网络中的函数包括传递函数、学习函数和性能函数。

传递函数通常使用S型对数或正切函数，由于本研究将输入输出数据都归一化处理到[-1,1]范围内，符合使用Sigmoid型正切函数对于数值区间的取值要求，所以传递函数选取隐层传递函数tansig和输出层传递函数logsig。采用带有动量梯度下降法作为网络的训练方法，学习函数采用learngdm函数，这个学习函数可以采用输入、误差、权重及阈值的学习率和动量常数来计算权重或者阈值的变化率，训练函数选择traingdm函数。

(4)初始化权重和阈值确定。采用遗传算法优化BP神经网络初始化权值和阈值的方法如下：

个体编码，生成初始的种群。采用实数编码方式对个体进行编码。编码串由隐层与输入层连接权值、输出层与隐层连接权值、隐层阈值、输出层阈值。将网络的权值和阈值按照一定的顺序级联起来，形成一个实数数组，作为遗传算法的一个染色体。编码的长度见公式(4)。在连接权值和阈值范围内，产生种群M个染色体构成初始群体。由于种群的规模对遗传算法的全局搜索性能有很大的影响，因此，种群的规模要根据具体的问题选取合适的数量。

$$S=n\times n_l+n_l\times m+n_l+m \qquad 公式(4)$$

其中，S为种群的规模，n_l为隐层节点个数，n为输入层节点个数，m为输出层节点个数。

适应度函数的设定。遗传算法进化搜索过程以适

应度函数为依据,利用种群中每个染色体的适应度值搜索,适应度值较高的个体遗传到下一代的概率较大。将适应度函数设定为 BP 神经网络误差 的倒数,当该适应度函数为最大值时 BP 神经网络的权重和阈值得到最优化,如公式(5)所示:

$$f(i) = \frac{1}{MSE_i} \quad 公式(5)$$

公式(5)中,$f(i)$表示第 i 条染色体的适应度值;MSE 为 BP 神经网络的预测输出与期望输出之间的误差平方和。

个体的选择。选择操作采用排序方法,按照个体适应度值的大小由小到大排列,最小适应度值的个体对应的序号为 1,最大适应度值的个体对应序号为 M。然后根据个体的适应度值的大小,按照适应度比例选择法计算个体的选择概率。概率值计算如公式(6)所示:

$$P_i = \frac{f_i}{\sum_{i=1}^{m} f_i} \quad 公式(6)$$

公式(6)中,f_i 为个体 i 的适应度值;m 为种群个体数目。

交叉操作和变异操作。交叉操作采用单点交叉,最优个体没有交叉操作,而是直接复制进入到下一代。对于其他个体,则使用交叉概率 p_c 表示对 2 个个体交叉操作,产生另外 2 个新个体的概率。同样,最优个体也没有进行变异操作,而是直接复制到下一代。变异操作采用均匀变异,对于其他的个体,则是用变异概率 p_m 进行变异操作,产生出另外新的个体。然后计算当前全体中每个染色体的适应度值,找出当前最优适应度值的个体,反复迭代,直到满足条件为止。

循环操作步骤(2)-(4),直到训练目标达到设置要求或者迭代的最大次数为止,获得 *BP* 神经网络的初始权值和阈值。

3.2.3 *GA-BP* 神经网络的训练过程　将 *GA-BP* 神经网络方法学习训练应用于社会化问答社区用户生成答案质量评价,就是将实际输出的答案质量评价等级值 y 和期望质量评价等级值 Y 值进行比较分析,如果实际输出答案质量评价等级值和期望答案质量评价等级值不相等,那么会根据相关误差计算公式得到误差,然后把误差信号按照原来的路径进行反向传输,利用输入不同的样本数据进行学习和训练分别得到输入层和隐层、隐层和输出层之间的权重系数,从而使得误差 *MSE* 值越来越小。一直到误差小于设定的阈值或最大训练次数,然后停止训练。*GA-BP* 神经网络经训练和学习后得到评价网络模型的权值和阈值、结构和隐层节点个数,形成社会化问答社区用户生成答案质量评价模型。然后,输入测试集样本数据,利用此评价模型自动化评价。输出层会输出实际效用值 y,将 y 利用函数 *postmnmx* 函数将其还原成真实值得到该答案质量评价结果,从而完成答案质量评价。

4 应用研究——以“知乎”网站为例

4.1 数据采集和预处理

本研究选取知乎网的问题“如何评价华为 Mate 10 & Mate 10 Pro?”下方的答案文本作为质量评价方法应用对象。该问题截至 2018 年 1 月 20 日拥有 494 个回答文本。采用 GooSeeker 软件采集该问题情境下问题提问者或浏览者、回答者、答案文本的相关数据。采用上述 3.1 小节的方法量化各个指标。指标量化过程中发现,由于知乎问答社区用户不存在等级和权威,本研究采用关注者数量量化,认为用户关注者数量越多,用户的权威性越高;知乎问答社区回答者用户也不存在最佳答案数量,采用回答者被知乎日报和知乎圆桌收录的问题数量量化最佳答案数量,认为回答者答案被知乎日报或者知乎圆桌收录,说明该答案具有权威性和代表性,可以认定为最佳答案。

由于本研究需要考虑不同的用户需求对答案文本质量的影响,所以选取 10 位 18-35 周岁的经常使用知乎 APP 的用户作为调研对象,编号为用户 1-10,从用户感知角度利用十分制的方法人工标注答案质量等级,没有确定的评价标准,仅凭用户主观判断标注答案质量等级。另外,由于本文仅选取一个提问问题进行应用研究,所以,用户提问数量为相同值,对于输出没有影响,可以不予考虑,只用于多个用户之间的比较分析。按照一般性的经验要求,神经网络模型构建时,样本选择需要符合二八定律,即训练样本数为总样本数的 80%,测试样本数为总样本数的 20%。所以,分别将编号前 400 的评论作为训练样本,编号 401-494 的 94 条评论作为测试样本。

4.2 答案质量评价方法应用分析

4.2.1 不同算法的比较分析　首先将上述采集到数据运用标准 BP 神经网络、SVM、最大熵、GA-BP 神经网络 4 种方法对比分析。采用 Matlab2015a 作为软件平台,利用神经网络工具箱函数、遗传算法工具箱编程实现 BP 神经网络和 GA-BP 神经网络效用评价方法的构建、训练和仿真,同样也实现 SVM、最大熵算法。选取文本特征维度、回答者特征维度、时效性维度的指标作

为基本特征(baseline),以用户1数据样本为基础针对基准特征进行测试。利用准确率 和平均误差值 来测量各类算法的准确性和性能。准确率 是指测试样本集中能够准确判断评价的样本数量所占的比例,当分类实际值与期望值差值绝对值控制在0.3以内可以认为准确,准确率 越高说明该方法判断的准确性越高;平均误差值 用每个测试样本误差绝对值之和求平均进行表示,平均误差值越小,表示模型的精确度和合理性越强。结果见表6。可以发现GA-BP神经神经网络算法的评价效果要好于其他分类算法,准确率较高,误差相对较低,可以应用于用户生成答案质量评价。

表6　不同分类算法的分类结果

评价方法	准确率 P	平均误差 M
标准 BP 神经网络	65.15%	0.66
最大熵	62.77%	0.76
SVM	63.83%	0.69
GA-BP	70.15%	0.58

4.2.2　GA-BP神经网络的应用研究　通过设置参数和函数构建基于GA-BP神经网络的社会化问答社区用户生成答案质量评价模型,分别对比分析基于基准特征加入用户特征维度、社会情感特征后对评价结果的影响。GA-BP神经网络采用3层的网络结构,由于不考虑用户提问问题数量指标特征,所以输入层的神经元个数为15。隐层节点数量采用实验试凑方法对隐层神经节点数量进行确定,发现当隐层节点个数为10时,MSE的值最小,所以将隐层节点个数设置为10。学习率 $\mu=0.01$,最大训练次数为100次,目标误差为0.01。将编号1－400的答案文本特征数据作为训练样本,将编号401－494的样本作为测试样本。利用遗传算法进行优化得到BP神经网络最优的初始值和阈值,设置定义遗传算法参数初始化种群数量为40、最大遗传代数MAXGEN＝80、采用实数编码染色体长度121、交叉概率为 px＝0.2、变异概率为 pm＝0.1。然后将遗传算法优化后得到的权值和阈值带入BP神经网络,重新进行训练,分别对10位用户编号401－494的测试样本进行评价验证。实验发现10位用户的训练样本采用GA-BP神经网络均在100步以内停止迭代达到目标误差0.01。以用户1为例,用户1的训练样本采用GA-BP神经网络方法,选取基准特征＋用户特征＋情感特征为输入,当遗传算法迭代代数在40次以内时寻到最优值,见图4;BP神经网络运行11次后停止迭代达到目标误差0.01,见图5。用户1的测试样本的期望值和实际输出值的结果,见图6。

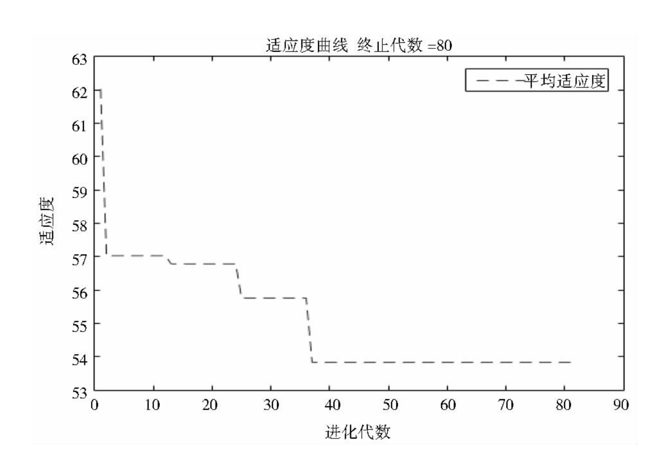


图4　遗传算法迭代次数

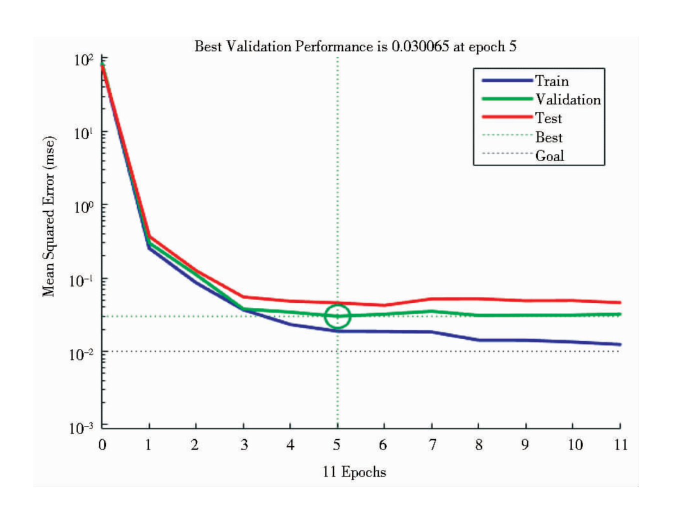


图5　GA-BP神经网络训练过程

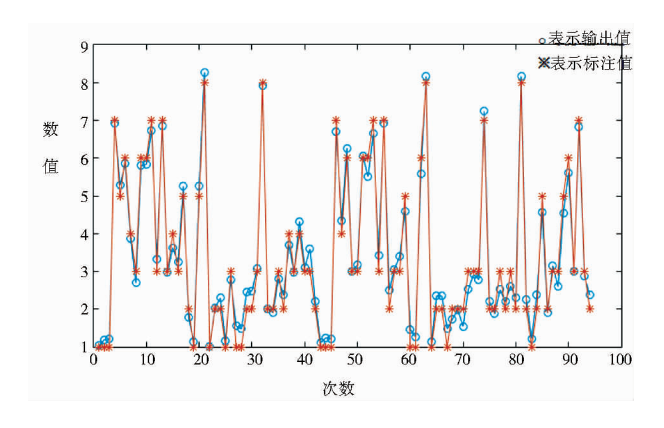


图6　用户1标注值与实际输出值的对比

分别基于GA-BP神经网络评价方法,当输入特征为基准特征、基准特征＋用户特征、基准特征＋社会情感特征、基准特征＋用户特征＋社会情感特征时用户1－10的测试样本的准确率 P 与平均误差 M,如表7所示:

表 7 GA-BP 神经网络模型效用评价准确率 *P* 和平均误差 *M*

用户编号	基准特征		基准特征 + 用户特征		基准特征 + 社会情感特征		基准特征 + 用户特征 + 社会情感特征	
	准确率 *P*	平均误差 *M*	准确率 *P*	平均误差 *M*	准确率 *P*	平均误差 *M*	准确率 *P*	平均误差 *M*
用户 1	69.15%	0.58	71.28%	0.64	72.34%	0.66	78.72%	0.43
用户 2	62.77%	0.65	64.89%	0.68	71.28%	0.62	72.34%	0.62
用户 3	71.28%	0.81	70.21%	0.75	68.09%	0.69	74.47%	0.70
用户 4	73.40%	0.89	72.34%	0.74	72.34%	0.63	77.66%	0.69
用户 5	75.53%	1.13	77.66%	0.82	79.79%	0.85	79.79%	0.56
用户 6	60.64%	0.85	65.96%	0.65	67.02%	0.61	72.34%	0.64
用户 7	62.77%	0.63	73.40%	0.53	71.28%	0.65	73.40%	0.63
用户 8	67.02%	0.61	65.96%	0.61	69.15%	0.59	75.53%	0.58
用户 9	69.15%	0.54	71.28%	0.57	76.60%	0.56	78.72%	0.52
用户 10	78.72%	0.58	79.79%	0.54	77.66%	0.57	80.85%	0.43

5 结果讨论与分析

通过上述研究发现：

(1)通过对比分析各类评价方法结果可以看出 GA-BP 神经网络方法能够应用于社会化问答社区用户生成答案质量评价，其评价的准确率虽然没有达到已有研究的最高准确率，但是当选取本文设计的特征时，明显高于 SVM 和最大熵方法的准确率。而且从图 5 可以看出 GA-BP 神经网络方法在迭代收敛速度方面明显高于标准 BP 神经网络，拥有较快的收敛速度，在 11 步以内就能够快速实现样本训练学习，而且实现了 100% 达到目标误差，不容易陷入局部最小值和无限循环中，能够快速构建社会化问答社区用户生成答案质量评价模型。因此，可以说明该方法能够应用于社会化问答社区用户生成答案质量评价，具有一定的合理性和科学性，能够进一步应用和推广。

(2)从表 5 可以看出当采用基准特征 + 用户特征、基准特征 + 社会情感特征时，虽然部分用户数据样本准确率 P 提升较少，但是平均误差 M 减少了很多，说明选取基准特征加用户特征时，可以使得评价值更接近目标值；当采用基准特征 + 用户特征 + 社会情感特征时准确率 P 明显提升，平均评价准确率 为 76.38%，具有较好的准确率，其平均误差 也较低，这说明 GA-BP 神经网络的质量评价值更加接近用户标注的真实值，表明该方法具有较强的仿真性和实用性，引入社会情感特征和用户特征后能够提高评价的准确率，笔者设计的评价指标体系具有一定的合理性和有效性。另外，从应用过程和结果可以看出基于 GA-BP 神经网络构建面向用户需求的答案质量评价方法，还能够根据不同用户信息需求和特点进行学习和训练，可以通过神经网络的训练学习，找寻输入和输出之间的内在联系，以权重的形式保存在神经网络中，不断自适应和调整，根据不同用户信息需求设计个性化的质量评价体系，从而增加评价模型的适应性和通用性。形成面向用户需求的个性化用户生成答案质量评价方法，具有一定的灵活性和个性化。

(3)对于社会化问答社区而言，保障平台用户生成内容质量和提供高质量的知识服务是推动平台发展的动力。社会化问答社区应该根据不同用户需求和特点对新生成的答案质量进行评价和筛选，可视化呈现高质量答案给用户，进而促进社区优质答案内容的传播。根据本文结论，社会化问答社区应当从答案文本特征、回答者特征、时效性、用户特征、社会情感特征等角度对优质答案内容进行挖掘和评价。可以采用机器学习中的人工神经网络模型(如 BP 神经网络)等方法进行评价和筛选优质内容，通过向用户推荐和呈现优质内容，控制和优化平台答案内容质量吸引新用户，也可以为老用户建立社区认同感，从而进一步促进社会化化问答社区可持续发展。

本研究为了解决社会化问答社区用户生成答案质量自动化评价问题，针对存在的评价指标体系不够全面、模糊性和缺乏个性化等问题，引入社会情感特征和用户特征维度，运用因子分析和结构方程实证构建用户生成答案质量评价指标体系。基于 GA-BP 神经网络模型设计了答案质量自动化评价方法。最后，选取知乎网站数据对用户生成答案质量评价指标体系和自动化评价方法进行应用研究。应用结果表明本研究构建的评价指标体系和评价方法具有一定的合理性和有效性。但是研究仍然存在一定的不足，首先应用研究样本选取具有一定的局限性，仅部分选取“知乎网站”的数据验证方法应用的有效性和合理性，话题内容也比较单一，没有进一步地将方法拓展到各个领域和类

型的问答社区。话题数据抽样方面存在局限,可能会导致研究结论的偏差。在后续的研究中将进一步加大应用研究对象的选取,扩大方法应用范围和领域。其次,仅从文本、回答者等外部特征层面选取和量化评价指标,没有深入到答案文本语义层面,从语义内容方面评价用户生成答案质量。后续的研究中需要结合语义网、机器学习等技术进一步加强对于用户生成答案质量语义层面评价研究。另外,情境也是用户评价和筛选答案质量的重要影响因素,后续的研究中将进一步探讨不同维度因素对答案质量的影响。

参考文献:

[1] KIM S, OH J S, OH S. Best-answer selection criteria in a social Q&A site from the user - oriented relevance perspective[J]. Proceedings of the Association for Information Science and Technology,2007,44(1):1－15.

[2] ISHIKAWA D, KANDO N, SAKAI T. What makes a good answer in community question answering? An analysis of assessors' criteria [EB/OL]. [2018－12－26]. https://www. researchgate. net/publication/228449185.

[3] OH S, WORRALL A, YI Y J. Quality evaluation of health answers in Yahoo! answers: a comparison between experts and users[J]. Proceedings of the Association for Information Science and Technology, 2011, 48(1):1－3.

[4] FICHMAN P. A comparative assessment of answer quality on four question answering sites[J]. Journal of information science, 2011, 37(5): 476－486.

[5] CHUA A Y K, BANERJEE S. So fast so good: an analysis of answer quality and answer speed in community question - answering sites[J]. Journal of the Association for Information Science and Technology, 2013, 64(10):2058－2068.

[6] 孙晓宁,赵宇翔,朱庆华. 基于SQA系统的社会化搜索答案质量评价指标构建[J]. 中国图书馆学报,2015,41(4):65－82.

[7] 李翔宇,陈琨,罗琳. FWG1法在社会化问答平台答案质量评测体系构建中的应用研究[J]. 图书情报工作,2016,60(1):74－82.

[8] 张煜轩. 基于外部线索的社会化问答平台答案信息质量感知研究[D]. 武汉:华中师范大学,2016.

[9] 姜雯,许鑫,武高峰. 附加情感特征的在线问答社区信息质量自动化评价[J]. 图书情报工作,2015,59(4):100－105.

[10] 袁红,张莹. 问答社区中询问回答的质量评价——基于百度知道与知乎的比较研究[J]. 数字图书馆论坛,2014(9):43－49.

[11] 孔维泽,刘奕群,张敏,等. 问答社区中回答质量的评价方法研究[J]. 中文信息学报,2011,25(1):3－8.

[12] 罗毅,曹倩. 基于RIPA方法的社会问答平台答案质量研究[J]. 图书情报工作,2015,59(3):126－133,25.

[13] 姜雯,许鑫. 在线问答社区信息质量评价研究综述[J]. 现代图书情报技术,2014(6):41－50.

[14] JEON J, CROFT W B, LEE J H, et al. A framework to predict the quality of answers with non-textual features[C]//Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2006: 228－235.

[15] SHAH C, POMERANTZ J. Evaluating and predicting answer quality in community QA[C]//Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval. New York: ACM,2010:411－418.

[16] 李晨,巢文涵,陈小明,等. 中文社区问答中问题答案质量评价和预测[J]. 计算机科学,2011,38(6):230－236.

[17] 王伟,冀宇强,王洪伟,等. 中文问答社区答案质量的评价研究:以知乎为例[J]. 图书情报工作,2017,61(22):36－44.

[18] 崔敏君,段利国,李爱萍. 多特征层次化答案质量评价方法研究[J]. 计算机科学,2016,43(1):94－97,102.

[19] 胡海峰. 用户生成答案质量评价中的特征表示及融合研究[D]. 哈尔滨:哈尔滨工业大学,2013.

[20] WANG R Y, STRONG D M. Beyond accuracy: what data quality means to data consumers[J]. Journal of management information systems, 1996, 12(4): 5－33.

[21] JOHN B M, CHUA A Y K, GOH D H L. What makes a high-quality user-generated answer? [J]. IEEE Internet computing, 2011, 15(1): 66－71.

[22] LIU B, FEMG J, LIU M, et al. Predicting the quality of user-generated answers using co-training in community-based question answering portals[J]. Pattern recognition letters, 2015,3(58): 29－34.

[23] 徐安滢,吉宗诚,王斌. 基于用户回答顺序的社区问答答案质量预测研究[J]. 中文信息学报,2017,31(2):132－138.

[24] HONAG L, LEE J T, SONG Y I, et al. A model for evaluating the quality of user-created documents[C]//Asia information retrieval symposium. Berlin: springer, 2008: 496－501.

[25] LIU Y, BIAN J, AGICHTEIN E. Predicting information seeker satisfaction in community question answering[C]//Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2008: 483－490.

[26] TIAN Q, ZHANG P, LI B. Towards predicting the best answers in community-based question-answering services[EB/OL]. [2018－12－26]. http://www. public. asu. edu/～bli24/Papers/ICWSM2013. pdf.

[27] 刘高军,马砚忠,段建勇. 社区问答系统中“问答对”的质量评价[J]. 北方工业大学学报,2012,24(3):31－36.

[28] 来社安,蔡中民. 基于相似度的问答社区问答质量评价方法[J]. 计算机应用与软件,2013,30(2):266－269.

[29] CAI Y, CHAKRAVARTY S. Answer quality prediction in Q/A social networks by leveraging temporal features[J]. International journal of next-generation computing, 2013, 4(1): 1－27.

[30] LI B, JIN T, LYU M R, et al. Analyzing and predicting question

quality in community question answering services[C]//Proceedings of the 21st international conference on World Wide Web. New York: ACM, 2012: 775 – 782.
[31] 袁健,刘瑜. 基于混合式的社区问答答案质量评价模型[J]. 计算机应用研究,2017,34(6):1708 – 1712.
[32] ANAND D, VAHAB F A. Predicting post importance in question answer forums based on topic-wise user expertise[C]//International conference on distributed computing and Internet technology. Berlin: Springer,2015: 365 – 376.
[33] ARAI K, HANDAYANI A N. Predicting quality of answer in collaborative Q/A community[J]. Society and culture, 2013, 2(3): 21 – 25.
[34] 吴明隆. 结构方程模型——AMOS 的操作与应用[M]. 重庆:重庆出版社,2009:52 – 53.
[35] 朱双东. 神经网络应用基础[M]. 沈阳:东北大学出版社,2000.
[36] TANG H, WU E X, MA Q Y, et al. MRI brain image segmentation by multi-resolution edge detection and region selection[J]. Computerized medical imaging and graphics, 2000, 24(6): 349 – 357.
[37] JEMEL S, HISSEL D, PERA M C, et al. On-board fuel cell power supply modeling on the basis of neural network methodology[J]. Journal of power sources, 2003, 124(2): 479 – 486.

作者贡献说明:
郭顺利:论文初稿写作,数据采集和处理;
张向先:论文框架制定,修改和终稿审定;
陶兴:数据采集和处理;
张莉曼:论文格式修改,中英文摘要写作。

Research on Automated Evaluation of User Generated Answer Quality in Social Question and Answer Community
——Taking "Zhihu" as an Example

Guo Shunli[1] Zhang Xiangxian[2] Tao Xing[2] Zhang Liman[2]
[1] Media College, Qufu Normal University, Rizhao 276826
[2] Management School Jilin University, Changchun 130022

Abstract: [**Purpose/significance**] The paper aims to build the social QA community users to generate the quality evaluation index system, achieve automatic evaluation and selection of answers to user needs, and improve the quality of the community QA community service. [**Method/process**] The introduction of social emotional features and user characteristics, and factor analysis and structural equation analysis are used to build an index system for evaluating the quality of user generated answers. Then, based on the GA-BP neural network model, the automatic evaluation method of the answer quality is designed. The application of the quality evaluation index system and automatic evaluation method of user generated answers is studied. [**Result/conclusion**] The evaluation index system consists of 5 dimensions, including the characteristics of the answer text, the characteristics of the respondent, the timeliness, the user characteristics and the social emotional characteristics. The experimental analysis shows that the method of automatic evaluation of the answer quality based on GA-BP neural network is more accurate and lower than other methods. It is feasible and effective, and can be further applied and popularized.

Keywords: social question and answer community user generated answers quality evaluation user demand